

**BROOKHAVEN NATIONAL LABORATORY
PROPOSAL INFORMATION QUESTIONNAIRE
LABORATORY DIRECTED RESEARCH AND DEVELOPMENT PROGRAM**

PRINCIPAL INVESTIGATOR Torre Wenaus

PHONE (631) 344-4755

DEPARTMENT/DIVISION Physics

DATE 04-25-2000

OTHER INVESTIGATORS Sasha Vanyashin (LBNL), Srinu Rajagopalan, Pavel Nevski, Victor Perevoztchikov, Yuri Fisyak

TITLE OF PROPOSAL **Algorithmic Virtual Data for HENP Computing Grids**

PROPOSAL TERM From (month/year) 10-01-2000 To (month/year) 10-01-2002

SUMMARY OF PROPOSAL

High Energy and Nuclear Physics (HENP) computing has a long history of innovation at the leading edge of computing, driven by the technical and scientific demands of successive generations of large scale HENP experiments. Current and next generation experiments stretch the bounds of today's computing technologies most particularly in a data analysis environment that is highly distributed and both data and computationally intensive. We propose to address a new computing challenge in a domain that has emerged as an important part of the solution to HENP's distributed large scale computing needs, Grid Computing. HENP physicists and developers are collaborating with IT researchers in the development of enabling tools and infrastructure to support Petabyte-scale computing in a highly distributed, multi-tier national and international architecture. Grid computing will employ a 'Virtual Data Grid' (VDG) providing a new degree of transparency in the delivery of data management and physics analysis capabilities to end users. While the VDG will operate in relatively conventional modes of mating existing data with processing services to fulfill user requests, it offers the potential for much more powerful and flexible analysis services involving the management and application of analysis algorithms – software, rather than static data – in the service of user requests. It is in managing such truly 'virtual' data that much of the most innovative development must occur. This is new ground, the payoff if successful being extremely powerful new capabilities not just in highly distributed analysis environments but in essentially all analysis work. It is in this area of providing capability for software algorithm based virtual data or 'algorithmic virtual data' (AVD) that this proposal has its focus.

While the Grid efforts being planned incorporate support for AVD, specific provision for AVD capability in application environments is where much of the complexity and challenge of AVD lies. This project will exploit the commonality in the software environments employed by HENP to develop an AVD capability broadly applicable within our community. We will apply our team's broad experience and innovative approaches in relevant HENP software domains to develop and deploy AVD in HENP in both non-Grid and Grid contexts. We will leverage and couple closely to BNL's existing roles in experiments and programs pursuing Grid Computing, including ATLAS and RHIC and BNL's large HENP Computing Facility. We will seek to build on the successful outcomes of this project to establish a long term leading role for BNL in Grid-enabled HENP software with support from the Grid and/or HENP communities. We seek support for two software developer FTEs for the two year duration of this project.

HUMAN SUBJECTS (Reference: DOE Order 1300.3)

Are human subjects involved from BNL or a collaborating institution? Y/N___N__

If YES, attach copy of the current Institutional Review Board
Approval and Informed Consent Form from BNL and/or
collaborating institution.

VERTEBRATE ANIMALS

Are vertebrate animals involved? Y/N___N__

If yes, has approval from BNL's Animal Care and Use
Committee been obtained?

Y/N_____

NEPA REVIEW

Are the activities proposed similar to those now carried out in the
department/division which have been previously reviewed for potential
environmental impacts and compliance with federal, state, local rules and
regulations, and BNL's Environment, Safety, and Health Standards.
(Therefore, if funded, proposed activities would require no additional
environmental evaluation.)

Y/N___Y___

If no, has a NEPA review been completed in accordance with
the Subject Area National Environmental Policy Act (NEPA) and Cultural
Resources Evaluation and the results documented?
Y/N_____

(Note: if a NEPA review has not been completed, submit a copy of the work proposal to
the BNL NEPA Coordinator for review. No work may commence until
the review is completed and documented.)

ES&H CONSIDERATIONS

Does the proposal provide sufficient funding for appropriate decommissioning
of the research space when the experiment is complete? Y/N___Y___

Is there an available waste disposal path for project wastes throughout the course of the
experiment? Y/N___Y___

Is funding available to properly dispose of project wastes throughout the
course of the experiment? Y/N___Y___

Can the proposed work be carried out within the existing safety envelope
of the facility (Facility Use Agreement, Nuclear Facility Authorization
Agreement, Accelerator Safety Envelope [ASE], etc.) in which it will be
performed? Y/N___Y___

If not, what has to be done to prepare the facility to accept the work

(modify the facility, revise the SAR/SAD, revise the Facility Use Agreement, etc.)
and how will the modifications be funded? Y/N_____

FUNDING REQUESTED [ATTACH A DETAILED BUDGET BREAKDOWN]

[Break down the funding by fiscal year and by the broad categories of labor, materials and supplies, travel (foreign & domestic), services and subcontracts.

LDRD funds cannot be used to purchase capital equipment. Indicate the intent to use collaborators and/or postdoctoral students, if applicable. Identify the various burdens applied, i.e., materials, organizational contracts. The Laboratory G&A should not be applied.]

POTENTIAL FUTURE FUNDING

Identify below the Agencies and the specific program/office which may be interested in supplying future funding. Give some indication of time frame.

We will seek to leverage external funding from the Grid Computing effort as it develops over the next few years. Our development in the near term of realized implementations of AVD, one of the most innovative and challenging areas in the Grid domain, will give us a base of experience and capability that positions us well for direct Grid project support.

APPROVALS

Department/Division Administrator
Department/Division Head
Cognizant Associate Director

BUDGET REQUEST BY FISCAL YEAR

COST ELEMENT in thousands	FISCAL YEAR _2001_____	FISCAL YEAR _2002_____
Labor *	160. (2 FTE)	168. (2 FTE)
Fringe	57.	60.
Total Labor	217.	228.
Organizational Burden @ _11.99___ %	26.	27.
Materials	10.	10.
Supplies	1.	1.
Travel	5.	5.
Services	5.	5.
Total MST	21.	21.
Materials Burden @ _6.5___%	1.4	1.4
Sub-contracts		
Contracts Burden @ _6.5___%		
Electric Power	1.	1.
CCD Charge		
Other (specify)		
TOTAL PROJECT COST	266.4	278.4
* Labor (indicate type of staff and level of effort)		
Physicist/Adv Computer Analyst	1.0 FTE	1.0 FTE
Advanced Computer Analyst	1.0 FTE	1.0 FTE
List all materials costing over \$5000.		

Algorithmic Virtual Data for HENP Computing Grids

PROPOSAL FOR LABORATORY DIRECTED RESEARCH AND DEVELOPMENT

Yuri Fisyak, Pavel Nevski, Victor Perevoztchikov, Srini Rajagopalan, Torre Wenaus
Physics Department, Brookhaven National Lab

Sasha Vanyashin
Lawrence Berkeley National Lab
Berkeley, CA

Abstract

High Energy and Nuclear Physics (HENP) computing has a long history of innovation at the leading edge of computing, driven by the technical and scientific demands of successive generations of large scale HENP experiments. Current and next generation experiments stretch the bounds of today's computing technologies most particularly in a data analysis environment that is highly distributed and both data and computationally intensive. We propose to address a new computing challenge in a domain that has emerged as an important part of the solution to HENP's distributed large scale computing needs, Grid Computing. HENP physicists and developers are collaborating with IT researchers in the development of enabling tools and infrastructure to support Petabyte-scale computing in a highly distributed, multi-tier national and even international architecture. Grid computing will employ a 'Virtual Data Grid' (VDG) providing a new degree of transparency in the delivery of data management and physics analysis capabilities to end users. While the VDG will operate in relatively conventional modes of mating existing data with processing services to fulfil user requests, it offers the potential for much more powerful and flexible analysis services involving the management and application of analysis algorithms – software, rather than just static data – in the service of user requests. It is in managing such truly 'virtual' data that much of the most innovative development must occur. This is new ground, the payoff if successful being extremely powerful new capabilities not just in highly distributed analysis environments but in essentially all analysis work. It is in this area of providing capability for software algorithm based virtual data or 'algorithmic virtual data' (AVD) that this proposal has its focus.

While the Grid efforts being planned incorporate support for AVD, specific provision for AVD capability in application environments is where much of the complexity and challenge of AVD lies. This project will exploit the commonality in the software environments employed by HENP to develop an AVD capability broadly applicable within our community. We will apply our team's broad experience and innovative approaches in relevant HENP software domains to develop and deploy AVD in HENP in both non-Grid and Grid contexts. We will leverage and couple closely to BNL's existing roles in experiments and programs pursuing Grid Computing, including ATLAS and RHIC. We will seek to build on the successful outcomes of this project to establish a long term leading role for BNL in Grid-enabled HENP software with support from the Grid and/or HENP project communities. We seek support for two software developer FTEs for the two year duration of this project.

Introduction

High Energy and Nuclear Physics (HENP) computing has a long history of innovation at the leading edge of computing, driven by the technical and scientific demands of successive generations of large scale HENP experiments. Current and next generation experiments stretch the bounds of today's computing technologies most particularly in a data analysis environment that is highly distributed and both data and computationally intensive. The large experiments at CERN's LHC, Fermilab's Tevatron and BNL's RHIC collider must supply a geographically highly dispersed physics community numbering in hundreds to thousands with effective means to analyze data volumes at scales of 100 Terabytes to 10 Petabytes with aggregate compute rates of up to 120 Tflops.

Many of the challenges associated with this environment are being actively addressed, with some projects now coming into fruition. Local examples include the Grand Challenge Architecture [1] for efficient large scale HENP data management developed by a LBNL-led collaboration and now deployed in RHIC's STAR experiment, and the Networked Object-based Environment for Analysis (NOVA) [2] developed at BNL (with LDRD support) and with component tools now deployed in the STAR and ATLAS experiments. These are however partial steps; both GCA and NOVA for example can live as components within a full scale distributed Petabyte-capable architecture, but providing that architecture remains to be done.

The leading edge nature of HENP's computing challenge is reflected in the attraction the problem holds for today's IT researchers. An important example of this is the GriPhyN [3] project, a collaboration between IT researchers and HENP and other experimenters from ATLAS, CMS, LIGO and SDSS. This NSF-driven project and other associated so-called 'Grid Computing' projects (including the Particle Physics Data Grid, PPDG [4]) have as their objective the development of the enabling tools and infrastructure to support Petascale distributed computing with application to (and beyond) the participating experiments. The exponential growth of networking capacities and technologies together with massive applications such as those in HENP are driving rapidly increasing attention to Grid Computing.

The GriPhyN project has identified the 'Virtual Data Grid' (VDG) as a useful unifying concept describing the new technologies required. A VDG is large in extent and scale, incorporating large numbers of resources on multiple regional, national and worldwide distance scales. It layers new services over the network proper which implement coordination of remote and possibly heterogeneous resources via standardized interfaces. And most importantly to its users, a VDG provides a new degree of transparency in the delivery of data handling and processing capabilities to end users, with the VDG providing a behind-the-scenes mapping of a user request into an efficient delivery mechanism. The VDG can impose a managed and optimized retrieval process taking account of policy and security constraints as well as the distribution of resources in the grid.

The most straightforward means of request management provided by a VDG is providing access to existing data, whether by delivering the data to the user or by performing the requested processing at the data locale. A more powerful (and technically challenging) service, however, will support requests resolvable also by providing for the generation of requested data (where possible) if doing so would provide more efficient (or the only) resolution. In this case, the data associated with a request does not exist; rather, software algorithms with associated prescriptions for configuration and input data sources drive its generation to resolve the request. It is in managing such truly 'virtual' data that much of the most innovative development must occur. This is new ground, the payoff if successful being extremely powerful new capabilities not just in highly distributed analysis environments but in essentially all analysis work.

It is in the area of providing capability for software algorithm based virtual data or 'algorithmic virtual data' (AVD) that this proposal has its focus.

Algorithmic Virtual Data for HENP

While major Grid Computing endeavors such as GriPhyN incorporate support for AVD in their programs, specific provision for AVD capability in application environments is where much of the complexity and challenge of AVD lies, and this domain extends beyond the scope of such projects into dependencies upon the software environment of the application domain. This project will exploit the commonality in the software architecture, organization, management and archival tools existing within HENP to develop an AVD capability broadly applicable within our community. While Grid architectures will provide the foundation and context for AVD, much room for innovation remains in closing the large gap between existing HENP computing environments and realized AVD capability.

Our implementation of AVD will assume the following characteristics – common to most current and future HENP experiments – are met the experiments' software. A less restrictive implementation would be possible but would require greater effort and could compromise the functionality.

- **Modular component based software architecture.** A modern OO architecture with algorithms implemented as object-based components with well-defined interfaces.
- **Run time component selection.** Shared libraries or an analogous mechanism support dynamic run-time specification of the program components.
- **Object based transient and persistent data models.** Data representations both in memory (transient) and storage based (persistent) are object based.

Elements of an algorithmic virtual data object can be summarized as follows.

- **Logical component name(s)** mapping onto the transient data model
- **The algorithms themselves**, of several forms; examples common in HENP which we will support are C++ shared libraries, Java classes, and perl script sets
- **Code archiving and management tools** managing the algorithms, including versioning.
- **A prescriptive generation procedure** for the data involving

-
- a set of one or more processing stages
 - software execution environment and component set for each stage
 - associated database requirements for each stage
 - configuration parameter sets for each stage
 - **Version information** for
 - data components
 - software components
 - database retrievals required in processing
 - configuration parameters required in processing
 - **Cost estimation** for the required processing.
 - **Supported execution environments** involving platform, OS, infrastructure type and version requirements

An outline of the steps involved in servicing AVD requests follows.

- **Virtual data references resolve to AVD**
- **Availability of data sources checked;** request extended to upstream source if necessary
- **Software build prescription retrieved**
- **Associated configuration parameters retrieved**
- **Database retrieval requirements resolved;** availability checked
- **Execution constraints resolved** (eg. platform, OS version)
- **Execution platform determined** (via Grid toolkit, in Grid based execution)
- **Estimation report on execution time delivered** to user if requested, based on cost model, execution platform, data source availability
- **On estimation accept, script builds for the target execution environment performed**
- **Fully configured job submitted to execution platform**
- **Data delivered to application** in the context of the experiment's object data model, either transiently for direct analysis or persistently to storage for later analysis

Some of the principal AVD implementation elements we will pursue are as follows.

- **Software component catalog** supporting multiple languages (as cited above) and version management
- **Software component archive** interfacing AVD to the archival tools used in HENP
- **Software component specification mechanisms.**
- **Prescriptive specifications for algorithmic data manipulations.** How the data is generated by software operating on input data (if any). The specification scheme will be capable of representing the data manipulations characteristic of HENP – event-based data processed in long and complex chains of algorithms with the data model 'view' of the data used by algorithms changing down the processing chain.
- **Procedural tools for job assembly and execution** adaptable to HENP experiments' execution environments
- **Management of distributed update and synchronization for software components.** AVD will function in the highly dynamic analysis environment, involving frequent updates

of software and data components, that prevails in HENP physics analysis environments.

Technologies and approach

Grid based computing takes advantage of the rapidly expanding capacity and capability of networked computing infrastructure in the internet era. In this project we will seek similar advantage on the software side, leveraging the rapid and continuing growth of the web and open source software. Dynamic web content in fact operates somewhat analogously to AVD, with requests in a highly distributed environment resolving to generation and presentation of information, and some tools heavily used in this domain we can effectively use. A principal example is the open source MySQL relational database heavily used as a back end database on web servers, and which has proven very successful in HENP applications developed by our team and others, for example in the NOVA project and the STAR experiment. In contrast to many HENP software development efforts which focus on in-house solutions or small market commercial tools closely focused on the problem domain, we will seek to apply very widely used and mature (or rapidly maturing) tools and technologies, developing the necessary adaptive layers and interfaces to map them onto our problem. We believe the advantage gained in leveraging such tools will more than balance the effort involved. We will pursue a component software approach in which standard components from the bazaar of the open source world support and complement components developed in-house. Specific tools we expect to employ in addition to **MySQL** are **perl** for scripting and operating system interaction; the **Apache** web server for web-based management, monitoring, and user interaction; the **PHP** dynamic web page language; **Java** (as well as C++) for network-intensive programming and possibly user interfaces; a host of **GNU** code management tools, **Linux** based services and many others.

Relation to other projects

We have discussed the complementarity of our HENP-directed AVD implementation goals to the major Grid projects. Other areas in which this project will complement the major Grid projects are in scope and timescale. Our scope is broader than the application of AVD in Grid environments. The promise of AVD lies not just in Grid-enabled distributed computing but in creating a much more powerful, transparent and flexible physics analysis environment whether the analysis is distributed or not. The software management infrastructure we will develop to provide the precisely defined prescription-based execution environments required for AVD can be applied outside the Grid context to provide physics analysis production and development in general with new levels of configurability and ease of use in specifying and managing precise and reproducible analysis software and execution environment configurations. These prescriptive analysis configuration tools will mesh well with the NOVA distributed analysis infrastructure developed in an earlier project, enabling us to integrate ease of use with (limited) distributed capability independent of the availability of an underlying Grid infrastructure. The existing BNL HENP computing facilities consisting of large scale commodity clusters provide an excellent testbed environment. This near term applicability of AVD enabling development complementing its

application in Grid environments is an important consideration in the timeline we propose for this project. We want to begin *now* because our developments will have real world application as soon as we can complete them, and the duration is a *short* two years because we want our deliverables to be Grid-ready (tested outside the Grid as well as in Grid prototyping) two years from now – when AVD support is scheduled to be available (GriPhyN timeline). During the initial years of the Grid projects while they are occupied with developing the infrastructure foundation for AVD, we will develop, test and be ready to deploy AVD implementations in HENP.

Leveraging of other work and programs

Two years ago some of us initiated an LDRD project to develop the NOVA (Networked Object-based Environment for Analysis) experiment-independent tool set for the realization of distributed physics analysis in the object oriented software environments now prevalent in HENP. This project is coming to a successful conclusion in FY00 with a tool suite implemented and components deployed in production as well as prototype environments in the STAR and ATLAS experiments at present. NOVA addressed principally the end user physics analysis environment with no AVD capability. Although the objectives of this project are distinct and of greater breadth, encompassing the provision for AVD throughout the full processing chain, several of the components developed within NOVA can find nearly direct application in this project. Most particularly, the cataloguing tools for data, software, and analysis job management, together with distributed production tools developed for large-scale data-intensive HENP production, will form the starting point for the advanced management and monitoring tools that must be developed for this project. In addition, NOVA's object request brokering tools and database interface tools will find application in interfacing our AVD capability to generic OO environments in HENP.

Developing AVD in a form that can efficiently and effectively mate to the computing models of HENP experiments demands a close understanding of the object data models used by the new generation of experiments targeted by this proposal. Our group brings many strengths in this area, with developers of the STAR, ATLAS and D0 data models among our participants. We also benefit from substantial experience in data storage, transport and management in HENP, particularly through the STAR experiment; our team encompasses the software management and much of the core development effort of STAR.

It is important at both the development and deployment stages that our project has a close connection to the HENP software environments we are targeting. At the development stage a close coupling will ensure that the 'real world' needs of HENP computing are reflected in our deliverables. At the deployment stage it will promote seeing our developments integrated into the production software of the experiments. Our group benefits from substantial, high level involvement in the current software efforts of two experiments, STAR and ATLAS. Two of us (SR, TW) hold management roles in U.S. ATLAS, including U.S. ATLAS Software Manager. Three of us (YF, PN, TW) hold management roles in STAR, including Software and Computing Leader.

One member of our team (SV) is a key participant in another major project that has developed enabling tools for large scale data management and analysis in HENP, the HENP Grand Challenge project. Sasha was formerly the principal NOVA developer, recently hired by LBNL (he is still BNL based) to integrate the Grand Challenge software into STAR. This LDRD project will enable Sasha to return to BNL employ and give our project the benefit of his experience and connections.

This project as an enabler for Grid based computing in HENP will benefit greatly from BNL's role as an emerging hub in the U.S. Grid fabric, and vice versa. In the RHIC program BNL is the computing hub at the center of a large worldwide community of active (lesser) RHIC computing sites. The value of Grid computing in this environment has long been recognized and the RHIC facility anticipates serving as a demonstration site for emerging Grid testbeds. In the U.S. ATLAS program BNL as the Tier 1 Regional Center is the principal U.S. computing center for ATLAS, charged with management of not only the Tier 1 Facility but the full U.S. Grid fabric incorporating Tier 2 centers around the country and the link to CERN. ATLAS and CMS are the principal drivers for Grid computing in HEP, seen as a key enabling technology for efficient and timely physics analysis, and BNL will play a central role. The presence of a major Grid center at BNL will enable close coupling and participation on the part of our project in leading edge Grid activities. In turn, the expertise in Grid-enabled software accrued by this project will both strengthen BNL's activities in Grid Computing and will aid in attracting direct Grid support for enabling software development as well as facilities.

Potential for external support

We will seek to leverage external funding from the general Grid effort as it develops over the next few years. Our development in the near term of realized implementations of AVD, one of the most innovative and challenging areas in the Grid domain, will give us a base of experience and capability that positions us well for direct Grid project support.

Work plan

In the first year of the project we will implement a prototype of the cataloguing tools; software specification, environment configuration and build tools; and interfaces to experiment computing environments and data models required for precisely determined software configuration and execution in response to a request. This will initially be applied outside the Grid context in prototype environments in one or more of the experiments to which our team has access (currently ATLAS, STAR).

In the second year we will iterate our design in the light of experience in non-Grid contexts (and, if Grid development timelines permit, also in Grid contexts in a limited way). We will complete the implementation of Grid-capable AVD following the architecture pursued by the Grid

community (GriPhyN, PPDG and collaborators) and demonstrate its application in an experimental setting (probably ATLAS).

At the conclusion of the project we will have a demonstrated, experiment independent AVD implementation integrated with the Grid architecture able to support HENP physics analysis. We will seek to establish it as a production tool in experimental settings. We will seek to leverage this work as a basis for continued Grid-enabled software participation with external funding.

Manpower and budget

To carry out our program of work we seek support for two FTEs for the next two years. One FTE will be used to recover Sasha Vanyashin from LBNL support and engage him 100% (or close to 100%) in this project as lead developer. Half an FTE will support major involvement from a software analyst or physicist closely coupled via the remainder of their time with another BNL activity with which we want this work to have close contact (either ATLAS, RHIC or the Grid Computing program of the HENP computing facility). They will participate in development with a focus on defining and developing interfaces to experiment environments and Grid infrastructure. This person is not yet identified, and may be a new person emerging from the growth taking place in some of these programs. In order to inject the experience base of all our team members in the design and development process the remaining half an FTE will support two concurrent 25% effort levels for development contributions from other members of our team, rotating among team members as appropriate to the work as the project progresses.

References

- [1] HENP Grand Challenge, <http://www-rnc.lbl.gov/GC/default.htm>
- [2] NOVA, <http://duvall.star.bnl.gov/nova>
- [3] GriPhyN, <http://www.phys.ufl.edu/~avery/griphyn/>
- [4] Particle Physics Data Grid, <http://www.cacr.caltech.edu/ppdg/>